British Journal of Pharmacology

British Journal of Pharmacology (2017) **174** 2174–2182 2174

Themed Section: Drug Metabolism and Antibiotic Resistance in Micro-organisms

RESEARCH PAPER

Xenobiotic-metabolizing enzymes in *Bacillus anthracis*: molecular and functional analysis of a truncated arylamine *N*-acetyltransferase isozyme

Correspondence Fernando Rodrigues-Lima, Sorbonne Paris Cité, Unité de Biologie Fonctionnelle et Adaptative, CNRS UMR 8251, Univ Paris Diderot, 75013, Paris, France. E-mail: fernando.rodrigues-lima@univ-paris-diderot.fr

Received 6 May 2016; **Revised** 1 September 2016; **Accepted** 26 September 2016

Xavier Kubiak[2], Romain Duval[1], Benjamin Pluvinage[4], Alain F Chaffotte[3], Jean-Marie Dupret[1] and Fernando Rodrigues-Lima[1]

[1]Sorbonne Paris Cité, Univ Paris Diderot, Unité BFA, CNRS UMR 8251, Paris, France, [2]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, [3]Unité de Résonance Magnétique Nucléaire des Biomolécules, Institut Pasteur, Paris, France, and [4]Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada

BACKGROUND AND PURPOSE

The arylamine *N*-acetyltransferases (NATs) are xenobiotic-metabolizing enzymes that play an important role in the detoxification and/or bioactivation of arylamine drugs and xenobiotics. In bacteria, NATs may contribute to the resistance against antibiotics such as isoniazid or sulfamides through their acetylation, which makes this enzyme family a possible drug target. *Bacillus anthracis*, a bacterial species of clinical significance, expresses three NAT isozymes with distinct structural and enzymatic properties, including an inactive isozyme ((BACAN)NAT3). (BACAN)NAT3 features both a non-canonical Glu residue in its catalytic triad and a truncated C-terminus domain. However, the role these unusual characteristics play in the lack of activity of the (BACAN)NAT3 isozyme remains unclear.

EXPERIMENTAL APPROACH

Protein engineering, recombinant expression, enzymatic analyses with aromatic amine substrates and phylogenetic analysis approaches were conducted.

KEY RESULTS

The deletion of guanine 580 (G580) in the *nat3* gene was shown to be responsible for the expression of a truncated (BACAN) NAT3 isozyme. Artificial re-introduction of G580 in the *nat3* gene led to a functional enzyme able to acetylate several arylamine drugs displaying structural characteristics comparable with its functional *Bacillus cereus* homologue ((BACCR)NAT3). Phylogenetic analysis of the *nat3* gene in the *B. cereus* group further indicated that *nat3* may constitute a pseudogene of the *B. anthracis* species.

CONCLUSION AND IMPLICATIONS

The existence of NATs with distinct properties and evolution in *Bacillus* species may account for their adaptation to their diverse chemical environments. A better understanding of these isozymes is of importance for their possible use as drug targets.

LINKED ARTICLES

This article is part of a themed section on Drug Metabolism and Antibiotic Resistance in Micro-organisms. To view the other articles in this section visit http://onlinelibrary.wiley.com/doi/10.1111/bph.v174.14/issuetoc

Abbreviations

AcCoA, acetyl CoA; (BACAN)NAT3, *Bacillus anthracis* NAT3; HDZ, hydralazine; INH, isoniazid; NAT, arylamine *N*-acetyltransferase; PAS, 4-aminosalicylic acid; PNPA, para-nitrophenylacetate; SMX, sulfamethoxazole; 2-AF, 2-aminofluorene[†]

[†]All NAT enzymes names follow the current NAT nomenclature (http://nat.mbg.duth.gr/).

DOI:10.1111/bph.13647

Table of Links

LIGANDS
Acetyl coenzyme A
Hydralazine

This Table lists key ligands in this article that are hyperlinked to corresponding entries in http://www.guidetopharmacology.org, the common portal for data from the IUPHAR/BPS Guide to PHARMACOLOGY (Southan *et al.*, 2016).

Introduction

Arylamine *N*-acetyltransferases (NATs, E.C.2.3.1.5) are a family of xenobiotic-metabolizing enzymes catalysing the acetyl CoA (AcCoA)-dependent acetylation of a wide range of aromatic amine chemicals, including drugs (Weber and Hein, 1985). Accumulating evidence suggests that NAT enzymes contribute to the adaptation of bacteria to their ecological niches through biotransformation of a variety of potential toxic aromatic compounds, including antibiotics (Payton *et al.*, 2001; Pluvinage *et al.*, 2007; Martins *et al.*, 2008; Kubiak *et al.*, 2012). In particular, NATs from *Mycobacterium tuberculosis* and *Bacillus anthracis* naturally increase their resistance to the anti-tubercular front-line drug isoniazid and the antibacterial drug sulfamethoxazole respectively (Payton *et al.*, 2001; Pluvinage *et al.*, 2007). In this respect, NATs represent interesting potential targets for the development of novel antibiotics, as exemplified by the development of piperidinol compounds inhibiting mycobacterial NATs and exerting anti-mycobacterial activity (Westwood *et al.*, 2010; Fullam *et al.*, 2011; Abuhammad *et al.*, 2012).

B. anthracis species, the agent responsible for anthrax disease, belongs to the larger *Bacillus cereus* group that comprises five other distinct but phylogenetically related species, namely, *B. cereus*, *B. thuringiensis*, *B. weihenstephanensis*, *B. mycoides* and *B. pseudomycoides* (Guinebretière *et al.*, 2008). The *B. anthracis* strain Sterne harbours three *nat* genes; two of which lead to fully functional NAT isoforms, whereas the *nat3* gene encodes an inactive C-terminally truncated protein (BACAN)NAT3 (Pluvinage *et al.*, 2007; Kubiak *et al.*, 2013b). Studies of *Salmonella typhimurium* (SALTY)NAT1 have shown that the C-terminal region was involved in controlling the arylamine-dependent hydrolysis of the acetyl CoA cofactor (Mushtaq *et al.*, 2002). In addition, (BACAN)NAT3 harbours a non-canonical Glu residue instead of an Asp at catalytic position 123 of the conventional Cys-His-Asp catalytic triad of NAT enzymes. We have recently reported that a homologue of (BACAN)NAT3 in *B. cereus* retains normal activity with a non-canonical Cys-His-Glu catalytic triad, although the same mutation abolishes the activity of the (HUMAN)NAT2 enzyme (Zang *et al.*, 2007; Kubiak *et al.*, 2013b). Interestingly, contrary to (BACAN)NAT3, its functional *B. cereus* homologue (BACCR)NAT3 is 263 amino acids long with a C-terminal domain displaying a 'classical' length (70 amino acids) (Kubiak *et al.*, 2013b). Several questions arise from these observations: (i) is the lack of NAT activity of the (BACAN)NAT3 isozyme due to the presence of the non-canonical catalytic triad or because of its short C-terminal domain (24 amino acids long)? (ii) What are the genetic causes leading to the truncation of the *nat3* gene? (iii) Does *nat3* constitute a pseudogene for the *B. anthracis* species?

In the present study, we identified a genetic change (frameshift mutation, c.580delG or p.(194 = fs*22)) in the *nat3* gene of *B. anthracis* that leads to a premature stop codon and to the expression of a (BACAN)NAT3 enzyme with a short C-terminal domain. We artificially restored the full-length enzyme [p.(194 = fs*68) or NAT3-fl] (263 amino acids with a C-terminus domain long of 70 amino acids), expressed it in *Escherichia coli* and purified it for biochemical and enzymatic studies. We found that it has comparable catalytic and structural properties with its *B. cereus* homologue (BACCR) NAT3, thus confirming the ability of other NAT isoforms bearing a non-canonical Cys-His-Glu catalytic triad to retain NAT activity. Importantly, it confirmed that the lack of NAT activity of (BACAN)NAT3 is due to its short C-terminus domain. Furthermore, phylogenetic analysis of the available genomes of the *B. cereus* group species indicates that the *nat3* gene constitutes a pseudogene in *B. anthracis* species, suggesting a divergent evolution and optimization of NAT isoform usage in *B. anthracis* species.

Methods

Cloning of B. anthracis nat3 ORF and generation of NAT3-fl enzyme

Cloning of the *nat3* open reading frame (ORF) (yvcN AB893_17270) coding for wild-type (BACAN)NAT3 protein has been previously reported by Pluvinage *et al.* (2007). We determined the c.580delG [or p.(194 = fs*22)] frameshift mutation responsible for the protein truncation by alignment with the orthologous *nat3* genes from *B. cereus* strain ATCC14579 previously characterized (Kubiak *et al.*, 2013b). Guanine 580 was subsequently reintroduced in the *nat3* ORF by three-step PCR mutagenesis (forward primer 5′-AGTGGATGAAGAAAAAGC-3′; reverse primer 5′-GCTTTTT-CTTCATCCACT-3′) coding for an artificial full-length enzyme (NAT3-fl). The primers used for amplification of the 5′- and 3′-ends of the *nat3* ORF and PCR programmes were as previously described by Pluvinage *et al.* (2007). PCR products were double digested using BamHI and XhoI (New England Biolabs, Evry, France), cloned into pET28(a) plasmid, verified by DNA sequencing and heterologously expressed as a hexa-histidine fusion protein in *E. coli* BL21.

Purification and SDS-PAGE analysis of recombinant enzymes

The expression and purification of (BACAN)NAT3, NAT3-fl and (BACCR)NAT3 were achieved as previously described (Pluvinage *et al.*, 2007; Kubiak *et al.*, 2013b). Purified recombinant proteins were analysed by Coomassie blue staining after SDS-PAGE and protein concentration determined by absorbance at 280 nm.

Activity assays and determination of apparent kinetic parameters

Activity rates and kinetic parameters were determined at steady state using the *para*-nitrophenylacetate (PNPA) assay (Cleland and Hengge, 1995). Typically, 0.5–1 μg enzyme and arylamine substrates were mixed in 25 mM Tris–HCl, pH 7.5, and 2 mM PNPA was added to initiate the reaction performed at 37°C. Product formation (PNP-) was measured by absorbance at 405 nm (ε_{PNP-} at 405 nm = 0.0035 μM·cm^{-1}), and data were normalized to the absorbance in the absence of arylamine substrate. The kinetic constants were obtained by non-linear regression of the average initial velocity (V_i) from three independent experiments (n = 3, with each n performed in technical triplicate) versus substrate concentration plot on the Michaelis–Menten equation (1), using Graphpad prism 2.0.

$$Vi = \frac{Vm^*[S]}{Km + [S]} \quad (1)$$

Circular dichroism experiments

Recombinant proteins were dialysed against 10 mM sodium phosphate, pH 7.5 buffer prior to circular dichroism (CD) spectra recoding (Aviv215 spectropolarimeter). Far-UV CD spectra were acquired between 180 and 260 nm with 435, 382 and 313 μg·mL^{-1} final concentrations of (BACAN) NAT3, NAT3-fl and (BACCR)NAT3, respectively, through a cylindrical cell with a 0.02 cm path length. The ellipticity signal was recorded each 0.5 nm for 1 s per step. Secondary structure proportions were deduced by deconvolution of the spectra against a 29-protein base (King and Johnson, 1999), using the CDPro package (http://lamar.colostate.edu/~sreeram/CDPro/main.html). Near-UV CD spectra were acquired between 250 and 320 nm with 2.17, 1.91 and 2.08-mg·mL^{-1} final concentrations of (BACAN)NAT3, NAT3-fl and (BACCR)NAT3 through a rectangular cell with 1 cm path length respectively. Average absorption spectra were obtained by five successive acquisitions and normalized with respect to the protein concentration and were expressed as the differential molar extinction coefficient Δε per residue and Δε per chain respectively.

Phylogenic analysis of the nat3 gene in B. cereus group

The *nat3* nucleotide sequence of the *B. anthracis* strain Sterne was used as a template to retrieve all akin *nat* gene sequences among both complete and draft genomes available for *B. cereus* group (taxid:86661). We used the BLASTn algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with all parameters set on default. A total of 373 nucleotide sequences were retrieved, including 82 from complete genomes and 291 from draft genomes. These were translated into amino acid sequences and aligned using MUSCLE. Phylogenetic tree reconstruction for sequences from the complete genomes was achieved with the web service www.phylogeny.fr (Dereeper *et al.*, 2010). Due to limitations in the number of sequences, we used MEGA6 (Tamura *et al.*, 2013) to build the phylogenetic tree of all of the *B. cereus* group *nat* sequences retrieved. Briefly, we used the neighbour joining analysis to determine phylogenic distances considering the Dayhoff model. The clades were statistically supported by 500 bootstrap replications.

Data and statistical analysis

All data are presented as means ± SEM. Statistical analysis of the activity measurements (n = 5 independent experiments, each conducted in triplicate) was performed by two-way ANOVA followed by Bonferroni (F achieved $P < 0.05$) correction using Prism 5.0 (GraphPad Software Inc., La Jolla, CA, USA), and Bartlett's test was used to assess homogeneity of variance of the data using R 3.3.1. Statistical significance is supported by $P < 0.05$. Data and statistical analysis comply with the recommendations on experimental design and analysis in pharmacology (Curtis *et al.*, 2015).

Chemical compounds

Unless stated otherwise, all chemical compounds were purchased from Sigma (Lyon, France).

Results

The *nat3* gene from *B. anthracis* strain Sterne spans 653 bp on the chromosome (locus tag BAS3292, from 3257790 to 3258580) and encodes a 217 amino acid protein that was previously shown to be inactive (Pluvinage *et al.*, 2007). In contrast, the orthologous 791 bp *nat3* gene from *B. cereus* strain ATCC 14579 (locus tag BC 3483, from 3438067 to 3438858) encodes a fully functional 263 amino acid enzyme (Kubiak *et al.*, 2013b). Neither of the two genes belongs to an operon according to the Prokaryotic Operon Prediction Database (Taboada *et al.*, 2012). In addition to their high sequence identity (91%), the overall organization of the loci 10 kb upstream and downstream, the *nat3* gene is very similar (Figure 1A), suggesting a speciation event, leading to divergent evolution. Alignment of the two genes (Figure 1B) reveals the deletion of a guanine nucleotide at position 580 in the *B. anthracis nat3* gene (c.580delG) leading to a premature stop codon at position 218 and a Glu194Lys at the site of mutation (p.(194 = fs*22)). On the amino acid level, the mutation removes most of (BACAN)NAT3 domain III (starting at Glu193), leaving intact domains I and II containing the non-canonical Cys69-His108-Asp123 catalytic residues and the two conserved NAT motifs found in (BACCR)NAT3. In contrast, 18 out of the 24 domain II residues remaining in (BACAN)NAT3 are different from (BACCR)NAT3 (Figure 1C). This previously unreported mutation is thus responsible for NAT3 isozyme of *B. anthracis* (BACAN)NAT3 being a short NAT protein with a short C-terminal domain compared with other characterized prokaryotic and eukaryotic NAT enzymes.

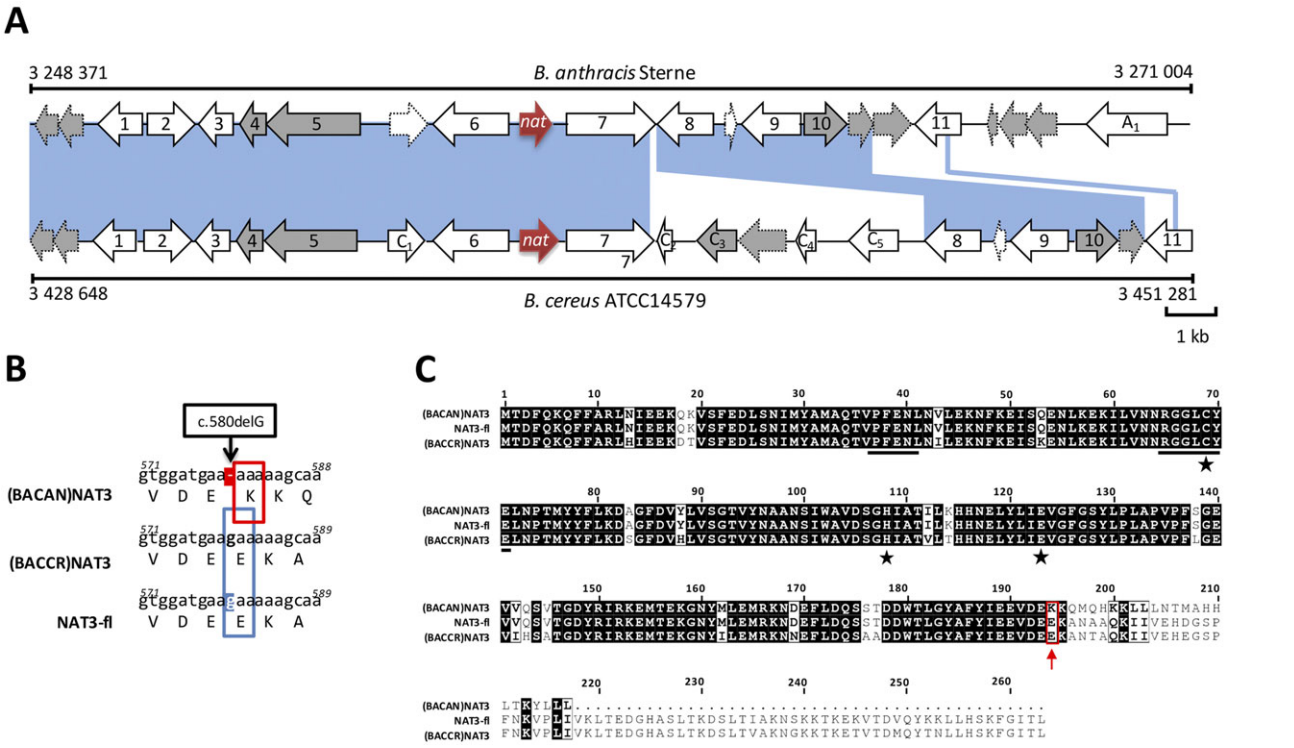

Figure 1

Locus organization and description of the *nat3* gene in *B. anthracis* (Sterne strain) and *B. cereus* (ATCC14579 strain). (A) Description of the *nat3* gene loci in *B. anthracis* Sterne and *B. cereus* ATCC14579 (−9 to +14 kb). Orthologous genes are highlighted in grey with the same numbers; different genes are numbered A_x and C_x for *B. anthracis* and *B. cereus* respectively. Genes coding for hypothetical proteins are shown as dashed arrows. Operons are highlighted in grey. The upstream loci are identical between the two species. Downstream, *B. cereus* ATCC14579 shows an insertion of five genes. The *nat3* genes are 651 and 791 bp in *B. anthracis* and *B. cereus* respectively. (B) In the *B. anthracis nat3* gene, the guanine 580 nucleotide (highlighted in red) is deleted and compared with *B. cereus*, leading to a frameshift (red frame), causing a premature stop codon at position 218 (underlined). The artificially restored full-length *B. anthracis* enzyme (NAT3-fl) was constructed by re-insertion of guanine 580 to restore the reading frame as in (BACCR)NAT3 (blue frame). (C) Multiple amino acid sequence alignment of (BACAN)NAT3, NAT3-fl and (BACCR)NAT3: the Glu194Lys mutation is highlighted by a red arrow, along with catalytic residues (star) and conserved NAT motifs (underlined).

To measure the impact of the truncated C-terminus on the enzyme's fold and activity, we artificially reintroduced the deleted guanine 580 in the *nat3* gene in order to compensate the frameshift and produce a 263 amino acid full-length version of (BACAN)NAT3, namely, NAT3-fl. NAT3-fl shows 90% identity with (BACCR)NAT3, with a fully restored domain III with almost identical amino acid composition (Figure 1C). The protein expressed well in heterologous expression system (*E. coli*) and was easily purified by affinity chromatography (6-His tag, data not shown).

The far-UV and near-UV spectra acquired in CD experiments indicate that NAT3-fl is properly folded and is very similar to its homologue (BACCR)NAT3 in both secondary structure and overall folding (Figure 2). (BACAN)NAT3 shows clear differences: its helices content is 6.9% and 9.3% lower than in NAT3-fl and (BACCR)NAT3, respectively, while the β-sheet content increase by 3.1–5.1% in (BACAN)NAT3 (Figure 2A). In addition, the four-times-higher root mean square deviation (r.m.s.d) value observed for (BACAN)NAT3 coincides with protein subpopulations that adopt different folds, in contrast to the structurally more homogenous (BACCR)NAT3 and NAT3-fl proteins (Khrapunov, 2009). Finally, the ellipticity signal in the absorbance zone of Trp (295 nm) and Trp/Tyr (280 nm) is significantly lower in (BACAN)NAT3, depicting a difference in the environment of these residues, likely a higher accessibility to the environment (Figure 2B).

We compared the acetyltransferase activity of (BACAN)NAT3 and NAT3-fl towards four prototypic arylamine substrates (Figure 3). As shown previously, (BACAN)NAT3 does not acetylate any of the arylamines (Pluvinage *et al.*, 2007, Figure S2). In contrast, NAT3-fl has specific activity rates of 526, 463, 416 and 852 nmol·min^{-1}·mg^{-1} for PAS, 2-aminofluorene, isoniazid and hydralazine substrates respectively. This first result demonstrates that the enzyme's lack of activity originates from its truncated C-terminus and confirms that, even bearing the non-canonical Glu residue in the catalytic triad, the NAT3-fl enzyme is active. When compared with the (BACCR)NAT3 homologue, NAT3-fl shows approximately half of the specific activity rate (2.2, 2.0, 1.9 and 1.6 times for PAS, 2-aminofluorene, isoniazid and hydralazine respectively) but keeps the same substrate specificity profile. As expected, no activity was measured for sulfamethoxazole with all enzymes (Pluvinage *et al.*, 2007; Kubiak *et al.*, 2013b). In accordance with specific activity rates, NAT3-fl has 1.6- to 1.8-times-lower k_{cat} values

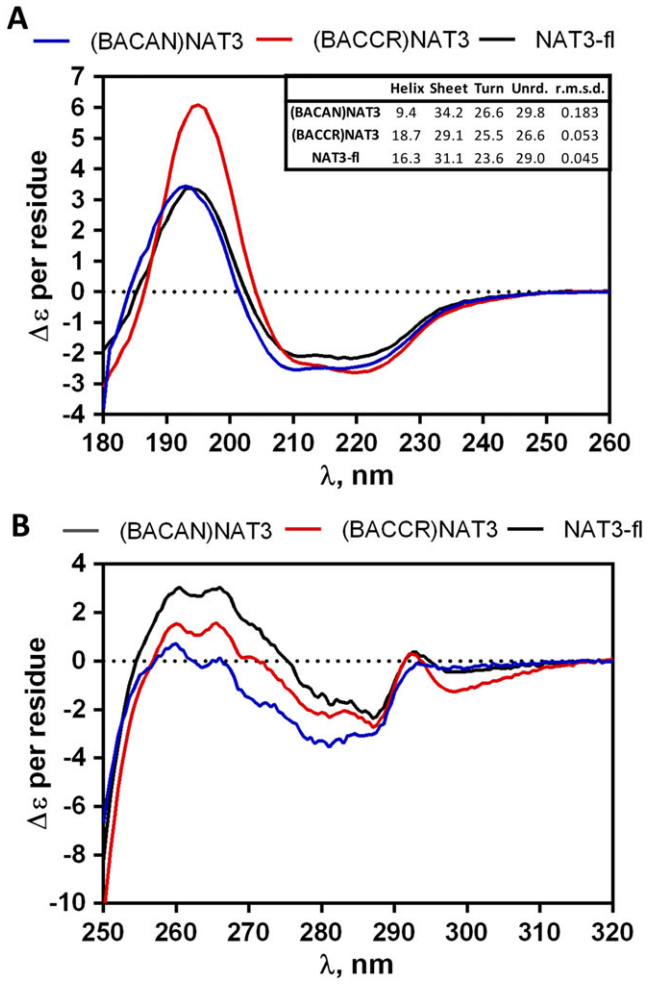

Figure 2

Circular dichroism (CD) spectra of (BACAN)NAT3, (BACCR)NAT3 and NAT3-fl. (A) Far-UV CD spectra (180–260 nm) of (BACAN)NAT3, (BACCR)NAT3 and NAT3-fl (blue, red and black lines respectively). Deconvolution of each spectrum indicates the relative proportion of secondary structures for each protein (inset). (B) Near-UV CD spectra (250–320 nm) of (BACAN)NAT3, (BACCR)NAT3 and NAT3-fl (the same colours as above).

compared with (BACCR)NAT3 (Figure 3). K_m^{app} values indicate that NAT3-fl has 3.3- and 1.6-times-higher affinity for 2-aminofluorene and hydralazine, respectively, compared with (BACCR)NAT3. In contrast, NAT3-fl has 1.2- and 1.8-times-lower affinity for PAS and isoniazid, respectively, compared with (BACCR)NAT3. Overall, these values are consistent with roughly two-times-lower catalytic efficiency (k_{cat}/K_m^{app}) for PAS and isoniazid, but a similar catalytic efficiency for hydralazine and a two-times-higher catalytic efficiency for 2-aminofluorene.

Finally, we investigated the phylogeny of the *nat3* gene in the *B. cereus* group to whom belongs *B. anthracis* species (Papazisi *et al.*, 2011; Reiter *et al.*, 2011). To do so, we screened all the available *B. cereus* group genomes available (complete and draft genomes) for nucleotide sequences similar to *nat3* from *B. anthracis* strain Sterne. All of the 373 sequences retrieved were hypothetic or established NAT sequences. Analysis of the complete genome sequences reveals that all *nat3* genes from *B. anthracis* strains are clustered in a unique clade and that they all encode a truncated isoform (Figure 4). In addition, the clade encompasses four *B. cereus* sequences and single *B. thuringiensis* sequences; however, these are full-length sequences like (BACCR)NAT3. The *nat* sequence from *B. thuringiensis BGSC 4AA1* is also a truncated isoform, but the stop codon is at position 150, suggesting that the mutation arises from a different evolutionary event. Among the draft genomes, truncated sequences are found only in *B. anthracis* strains, with the sole exception of *B. cereus SJ1* that clusters with it with significant probabilities (data not shown). All sequences retrieved possess the non-canonical Glu residue. The truncated (BACAN)NAT3 form is found in 80% of *B. anthracis* sequences (80 sequences out of 103 *nat* sequences from *B. anthracis* genomes). Together, these results suggest that the *nat3* gene from *B. anthracis* species has evolved from a common ancestor.

Discussion

In the present paper, we report the functional, structure-related and phylogenic characterization of the truncated (BACAN)NAT3 isoform of *B. anthracis* strain Sterne and an artificially restored full enzyme (NAT3-fl), based on the homologous *B. cereus* (BACCR)NAT3 enzyme. The *nat3* gene sequence analysis reveals that a G580 mutation is responsible for the truncation of the 46 C-terminal amino acids of (BACAN)NAT3 and that is bears the same non-canonical Cys-His-Glu non-canonical catalytic triad as (BACCR)NAT3.

(BACAN)NAT3 was described as unable to acetylate prototypic NAT substrates, but it remained non-elucidated whether it was because of its truncated C-terminal domain (domain III) or because of the non-conserved Glu residue in the catalytic triad (Pluvinage *et al.*, 2007). Indeed, the C-terminal region of NAT enzymes is known to be important to control the substrate-dependent AcCoA hydrolysis, and we have recently reported that the homologous (BACCR)NAT3 enzyme carrying the non-canonical Cys-His-Glu catalytic triad was functional (Kubiak *et al.*, 2013b).

CD data, along with the expression and purification of a soluble protein, indicate that (BACAN)NAT3 is properly folded, confirming that domains I and II (residues 1–192) are folding independently from domain III. This is consistent with previous observations on a 204-residue fragment of (HUMAN)NAT1 and a 196 amino acid construct of (SALTY)NAT1 that were soluble when expressed in *E. coli* (Sinclair and Sim, 1997; Mushtaq *et al.*, 2002). In light of the very high sequence similarity between NAT3-fl and (BACCR)NAT3, we can assume that both proteins adopt the same fold (Pluvinage *et al.*, 2011; Kubiak *et al.*, 2013b). The (BACCR)NAT3 structure also reveals that Glu194 side chain does not interact with any other residue in the protein (Figure S1), suggesting that all structural and functional effects observed stem the C-terminal region deletion.

Enzymatic characterization of (BACAN)NAT3, NAT3-fl and (BACCR)NAT3 reveal that NAT3-fl possesses catalytic properties similar to those of its *B. cereus* homologue, making NAT3-fl the second NAT isoform to be reported functional with a Cys-His-Glu catalytic triad (Kubiak *et al.*,

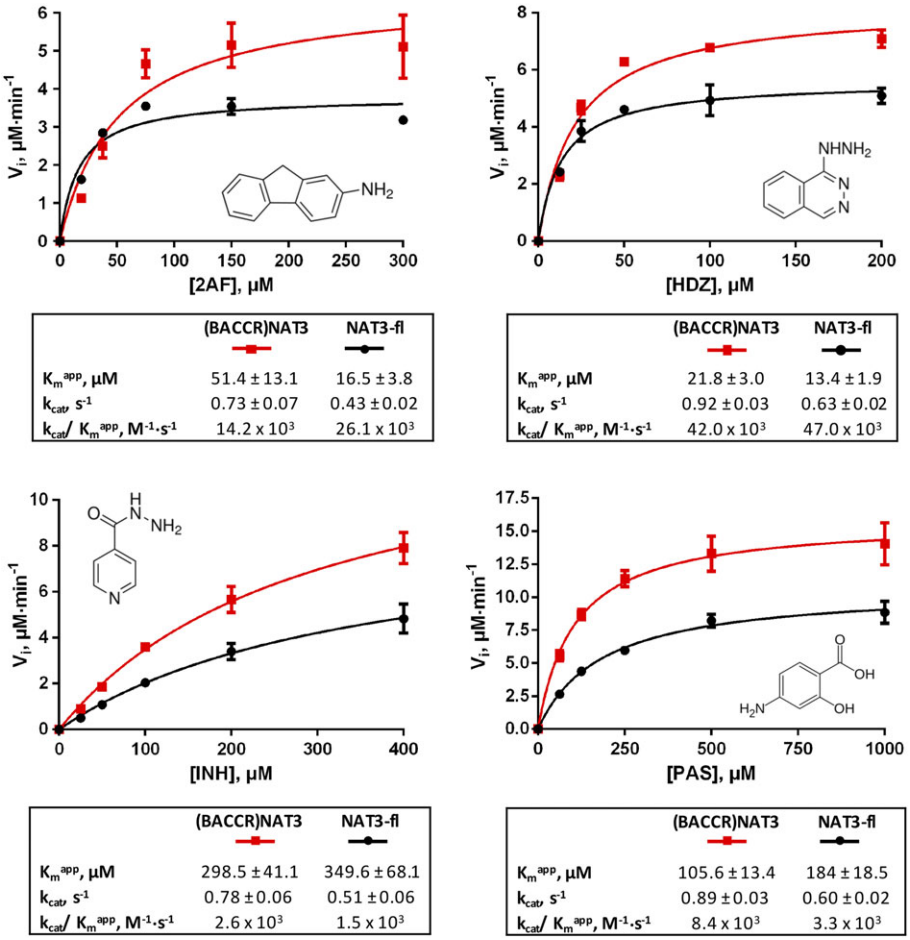

Figure 3

Comparison of (BACCR)NAT3 and NAT3-fl kinetic parameters. Michaelis–Menten saturation curves for 2-aminofluorene (2AF), hydralazine (HDZ), isoniazid (INH) and PAS in the presence of 2 mM PNPA and 0.5–1 µg enzyme. Kinetic parameters K_m^{app}, k_{cat} and k_{cat}/K_m^{app} indicated in insets. Average data from three independent experiments (each performed in technical triplicates; values shown as mean ± SEM) were fitted to the Michaelis–Menten equation (1) after normalization with PNPA hydrolysis rate in the absence of enzyme.

2013b). The slight differences in activity and kinetic parameters are rather standard in the family of NAT enzymes even for very closely related isoforms, notably as demonstrated for the three *Legionella pneumophila* NAT variants that belonged to the same species and have only few point mutations (Kubiak *et al.*, 2012). The contribution of the C-terminus truncation to the loss of activity of (BACAN)NAT3 is further supported by the absence of AcCoA hydrolase activity. Indeed, the *B. anthracis* (BACAN)NAT2 isoform (255 amino acids) possesses AcCoA hydrolase activity, although it is only 8 and 24 residues shorter than NAT3-fl and (BACAN)NAT1, respectively, both of which showing no AcCoA hydrolase activity (Pluvinage *et al.*, 2007). Similarly, studies on a 270 construct of (SALTY)NAT1 showed that the nine C-terminal residues were essential to control the arylamine-dependent AcCoA hydrolysis (Mushtaq *et al.*, 2002). Thus, our results, altogether with previous studies, suggest that (BACAN)NAT3 does not bind the AcCoA cofactor.

Co-crystallization experiments with CoA demonstrated that domain III of (BACAN)NAT1 and (HUMAN)NAT2 contains nine and seven out of 14 residues involved in the binding of AcCoA respectively (Wu *et al.*, 2007; Pluvinage *et al.*, 2011). In light of the 90% amino acid sequence identity and similarity of CD spectra, we assume that (BACCR)NAT3 is a good model to substantiate the effects of the G580 mutation on binding of AcCoA in (BACAN)NAT3. Superposition of the structures of the highly similar (BACCR)NAT3 and (BACAN)NAT1 bound to CoA (r.m.s.d = 0.75 Å) reveals that eight out of 14 residues interacting with AcCoA are absent in (BACAN)NAT3, removing all the interactions with the PPi-3′ ADP phosphate part of the cofactor, which is likely to impair its binding (Figure S1). Intriguingly, deletion of domain III in both (HUMAN)NAT1 and (SALTY)NAT1 was compatible with hydrolysis and thus binding of AcCoA (Sinclair and Sim, 1997; Mushtaq *et al.*, 2002). Although neither (HUMAN)NAT1 nor (SALTY)NAT1 was co-crystallized with CoA (Sinclair *et al.*, 2000; Wu *et al.*, 2007), the (HUMAN)NAT2 (an r.m.s.d of 0.7 Å with HUMAN(NAT1)) structure bound to CoA showed that residues interacting with the PPi-3′ ADP phosphate part of CoA are spread between domains II and III and that CoA has a significantly different orientation compared with (BACAN)NAT1 (Wu

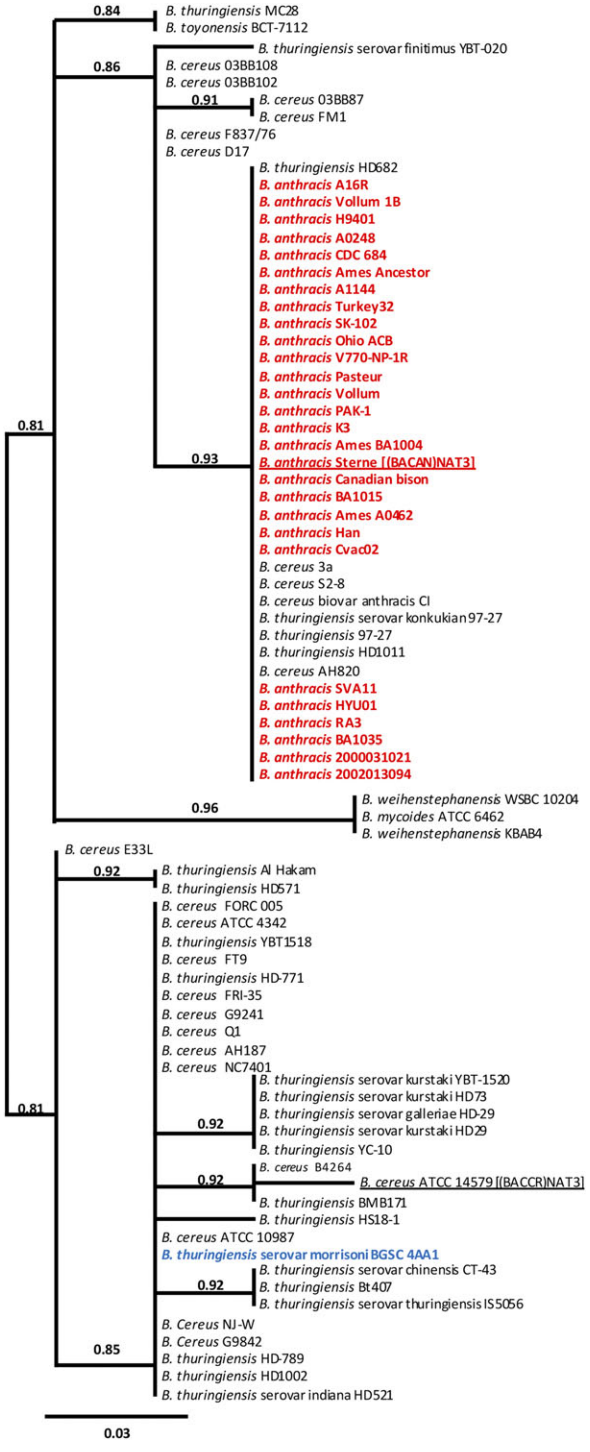

Figure 4

Phylogenic analysis of the *nat3* gene in *B. cereus* group. NAT amino acid sequences were searched for amongst all complete *B. cereus* group genomes using the *nat3* nucleotide sequence from *B. anthracis* (Sterne strain). Phylogenic relationships were inferred after sequence alignment with MUSCLE and tree building automatically within the Phylogeny.fr website with default parameters. Statistics supporting the clades are indicated in bold numbers. Truncated sequences are highlighted in bold red and the single non-*anthracis* truncated sequence in blue. Sequences corresponding to (BACAN)NAT3 and (BACCR)NAT3 are underlined.

et al., 2007, Figure S1). Thus, (HUMAN)NAT1 might still retain sufficient interaction with most of CoA to bind the cofactor. Owing to the large variety in the modes of binding of CoA within NAT enzymes, the domain III deletion might have different effects depending on the enzyme (Fullam *et al.*, 2008; Sim *et al.*, 2008b; Kubiak *et al.*, 2013a; Xu *et al.*, 2015).

Our phylogenic analysis of the *nat3* gene among the *B. cereus* group, together with the functional and structure-related data, supports the existence of an active ancestral protein that evolved toward an inactive form. Although (BACAN)NAT3 protein is not functional, mRNA expression of the *nat3* gene in *B. anthracis* strain Sterne has been shown (Pluvinage *et al.*, 2007). This is somewhat surprising as bacteria tend to eliminate unnecessary genes (such as an NAT enzyme with no AcCoA hydrolase or acetyltransferase activity) but can easily be accounted for ongoing evolution in the species. The phylogenetic relationship between *bacilli* has been a debate for decades whether they represent distinct species, but comparative genomic analyses have shown that *B. anthracis* has emerged as a distinct lineage within the *B. cereus* group through different genomic events such as gain, loss, duplication, internal deletion and lateral transfer (Papazisi *et al.*, 2011; Reiter *et al.*, 2011). The general locus organization around the two orthologous *nat3* genes in *B. cereus* ATCC14579 and *B. anthracis* Sterne is consistent with the evolution of *B. anthracis* from a common *B. cereus* ancestor or as a lineage of *B. cereus* (Papazisi *et al.*, 2011; Reiter *et al.*, 2011) and supports our observations (Figure 1). Surprisingly, all *B. anthracis nat* sequences segregate apart from the clade containing the *B. cereus nat3* gene, showing no correlation between the truncation of the NAT3 enzyme and the presence of the non-canonical Glu catalytic residue (Sim *et al.*, 2008a). Thus, the most likely scenario to account for a truncated NAT3 isoform is the mutation of *nat3* gene in a common ancestor that has been inherited in all descendant strains by a lack of selection pressure, probably because the NAT3 isoform has functional properties (e.g. substrate specificity) overlapping with those of the *B. anthracis* NAT1 and NAT2 isoforms (lack of selection pressure).

Interestingly, the emergence and evolution of *B. anthracis* by reduction/loss of genes coding for enzymes involved in metabolism (degradation of amines, biosynthesis of secondary metabolites, etc.) have been reported (Papazisi *et al.*, 2011). The persistence of a truncated non-functional gene in most *B. anthracis* species suggests that *nat3* represents a pseudogene for *B. anthracis*. The variation of *nat* genes in bacteria is poorly documented. The existence of a few polymorphisms in *Mycobacteria* and of enzyme variants in strains of *Legionella pneumophila* has been reported (Coelho *et al.*, 2011; Kubiak *et al.*, 2012). However, this is the first time to our knowledge that an NAT pseudogene in bacteria has been reported.

Acknowledgements

This work was supported by Université Paris Diderot – Paris 7, Délégation Générale de l'Armement (DGA), the Institut

Pasteur and the Centre National de la Recherche Scientifique (CNRS). X.K. was supported by a fellowship from the Université Paris Diderot – Paris7. R.D. is supported by a DIM grant from Région Ile-de-France. B.P. was supported by a fellowship from the DGA. We acknowledge the plateform "Biophysique des macromolécules et de leurs intéractions" (Institut Pasteur) for provision of CD facilities.

Author contributions

X.K. and F.R.L. participated in research design. X.K., R.D., B.P. and A.F.C. conducted experiments. X.K., R.D., B.P., A.F.C., J.M.D. and F.R.L. performed data analysis. X.K. and F.R.L. wrote or contributed to the writing of the manuscript.

Conflict of interest

The authors declare no conflicts of interest.

Declaration of transparency and scientific rigour

This Declaration acknowledges that this paper adheres to the principles for transparent reporting and scientific rigour of preclinical research recommended by funding agencies, publishers and other organisations engaged with supporting research.

References

Abuhammad A, Fullam E, Lowe ED, Staunton D, Kawamura A, Westwood IM *et al.* (2012). Piperidinols that show anti-tubercular activity as inhibitors of arylamine N-acetyltransferase: an essential enzyme for mycobacterial survival inside macrophages. PLoS One 7: e52790.

Cleland WW, Hengge AC (1995). Mechanisms of phosphoryl and acyl transfer. FASEB J 9: 1585–1594.

Coelho MB, Costa ERD, Vasconcellos SEG, Linck N, Ramos RM, Amorim HL *et al.* (2011). Sequence and structural characterization of tbnat gene in isoniazid-resistant *Mycobacterium tuberculosis*: identification of new mutations. Mutat Res 712: 33–39.

Curtis MJ, Bond RA, Spina D, Ahluwalia A, Alexander SP, Giembycz MA *et al.* (2015). Experimental design and analysis and their reporting: new guidance for publication in BJP. Br J Pharmacol 172: 3461–3471.

Dereeper A, Audic S, Claverie JM, Blanc G (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol 10: 8.

Fullam E, Abuhammad A, Wilson DL, Anderton MC, Davies SG, Russell AJ *et al.* (2011). Analysis of β-amino alcohols as inhibitors of the potential anti-tubercular target N-acetyltransferase. Bioorg Med Chem Lett 21: 1185–1190.

Fullam E, Westwood IM, Anderton MC, Lowe ED, Sim E, Noble MEM (2008). Divergence of cofactor recognition across evolution: coenzyme A binding in a prokaryotic arylamine N-acetyltransferase. J Mol Biol 375: 178–191.

Guinebretière M-H, Thompson FL, Sorokin A, Normand P, Dawyndt P, Ehling-Schulz M *et al.* (2008). Ecological diversification in the *Bacillus cereus* group. Environ Microbiol 10: 851–865.

Khrapunov S (2009). Circular dichroism spectroscopy has intrinsic limitations for protein secondary structure analysis. Anal Biochem 389: 174–176.

King SM, Johnson WC (1999). Assigning secondary structure from protein coordinate data. Proteins 35: 313–320.

Kubiak X, Dairou J, Dupret J-M, Rodrigues-Lima F (2013a). Crystal structure of arylamine N-acetyltransferases: insights into the mechanisms of action and substrate selectivity. Expert Opin Drug Metab Toxicol 9: 349–362.

Kubiak X, Dervins-Ravault D, Pluvinage B, Chaffotte AF, Gomez-Valero L, Dairou J *et al.* (2012). Characterization of an acetyltransferase that detoxifies aromatic chemicals in *Legionella pneumophila*. Biochem J 445: 219–228.

Kubiak X, Li de la Sierra-Gallay I, Chaffotte AF, Pluvinage B, Weber P, Haouz A *et al.* (2013b). Structural and biochemical characterization of an active arylamine N-acetyltransferase possessing a non-canonical Cys-His-Glu catalytic triad. J Biol Chem 288: 22493–22505.

Martins M, Pluvinage B, Li de la Sierra-Gallay I, Barbault F, Dairou J, Dupret J-M *et al.* (2008). Functional and structural characterization of the arylamine N-acetyltransferase from the opportunistic pathogen *Nocardia farcinica*. J Mol Biol 383: 549–560.

Mushtaq A, Payton M, Sim E (2002). The COOH terminus of arylamine N-acetyltransferase from *Salmonella typhimurium* controls enzymic activity. J Biol Chem 277: 12175–12181.

Papazisi L, Rasko DA, Ratnayake S, Bock GR, Remortel BG, Appalla L *et al.* (2011). Investigating the genome diversity of *B. cereus* and evolutionary aspects of *B. anthracis* emergence. Genomics 98: 26–39.

Payton M, Gifford C, Schartau P, Hagemeier C, Mushtaq A, Lucas S *et al.* (2001). Evidence towards the role of arylamine N-acetyltransferase in *Mycobacterium smegmatis* and development of a specific antiserum against the homologous enzyme of *Mycobacterium tuberculosis*. Microbiology 147: 3295–3302.

Pluvinage B, Dairou J, Possot OM, Martins M, Fouet A, Dupret J-M *et al.* (2007). Cloning and molecular characterization of three arylamine N-acetyltransferase genes from *Bacillus anthracis*: identification of unusual enzymatic properties and their contribution to sulfamethoxazole resistance. Biochemistry 46: 7069–7078.

Pluvinage B, Li de la Sierra-Gallay I, Kubiak X, Xu X, Dairou J, Dupret J-M *et al.* (2011). The Bacillus anthracis arylamine N-acetyltransferase ((BACAN)NAT1) that inactivates sulfamethoxazole, reveals unusual structural features compared with the other NAT isoenzymes. FEBS Lett 585: 3947–3952.

Reiter L, Tourasse NJ, Fouet A, Loll R, Davison S, Økstad OA *et al.* (2011). Evolutionary history and functional characterization of three large genes involved in sporulation in *Bacillus cereus* group bacteria. J Bacteriol 193: 5420–5430.

Sim E, Lack N, Wang C-J, Long H, Westwood I, Fullam E *et al.* (2008a). Arylamine N-acetyltransferases: structural and functional implications of polymorphisms. Toxicology 254: 170–183.

Sim E, Walters K, Boukouvala S (2008b). Arylamine N-acetyltransferases: from structure to function. Drug Metab Rev 40: 479–510.

Sinclair J, Sim E (1997). A fragment consisting of the first 204 amino-terminal amino acids of human arylamine N-acetyltransferase one (NAT1) and the first transacetylation step of catalysis. Biochem Pharmacol 53: 11–16.

Sinclair JC, Sandy J, Delgoda R, Sim E, Noble ME (2000). Structure of arylamine N-acetyltransferase reveals a catalytic triad. Nat Struct Biol 7: 560–564.

Southan C, Sharman JL, Benson HE, Faccenda E, Pawson AJ, Alexander SPH *et al.* (2016). The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. Nucleic Acids Res 44: 1054–1068.

Taboada B, Ciria R, Martinez-Guerrero CE, Merino E (2012). ProOpDB: Prokaryotic Operon DataBase. Nucleic Acids Res 40: D627–D631.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30: 2725–2729.

Weber WW, Hein DW (1985). N-acetylation pharmacogenetics. Pharmacol Rev 37: 25–79.

Westwood IM, Bhakta S, Russell AJ, Fullam E, Anderton MC, Kawamura A *et al.* (2010). Identification of arylamine N-acetyltransferase inhibitors as an approach towards novel anti-tuberculars. Protein Cell 1: 82–95.

Wu H, Dombrovsky L, Tempel W, Martin F, Loppnau P, Goodfellow GH *et al.* (2007). Structural basis of substrate-binding specificity of human arylamine N-acetyltransferases. J Biol Chem 282: 30189–30197.

Xu X, Li de la Sierra-Gallay I, Kubiak X, Duval R, Chaffotte AF, Dupret JM *et al.* (2015). Insight into cofactor recognition in arylamine N-acetyltransferase enzymes: structure of *Mesorhizobium loti* arylamine N-acetyltransferase in complex with coenzyme A. Acta Crystallogr D Biol Crystallogr 71: 266–273.

Zang Y, Zhao S, Doll MA, Christopher States J, Hein DW (2007). Functional characterization of the A411T (L137F) and G364A (D122N) genetic polymorphisms in human N-acetyltransferase 2. Pharmacogenet Genomics 17: 37–45.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

http://dx.doi.org/10.1111/bph.13647

Figure S1 (BACCR)NAT3-based model of the effect of G580 mutation in (BACAN)NAT3. Residue 194 affected by the G580 deletion and the C-terminal region of (BACAN)NAT3 are shown in red sticks and cartoon, respectively, on the (BACCR)NAT3 structure (PDB ID 4DMO). C-terminal domain absent in (BACAN)NAT3 is coloured in pink. (BACCR)NAT3 was superimposed with the CoA-bound (BACAN)NAT1 structure (PDB ID 3LNB): residues interacting with CoA (green sticks) that are absent or modified in (BACAN)NAT3 are highlighted in blue sticks. CoA from the CoA-bound (HUMAN)NAT2 structure (PDB ID 2PFR) is shown in black sticks.
Figure S2 Specific activity rates of (BACAN)NAT3, NAT3-fl and (BACCR)NAT3. PNP- product formation was followed during stationary phase at 405 nm, with each enzyme in presence of 2 mM PNPA and 500 μM substrate for each enzyme, was followed during stationary phase at 405 nm for using the PNPA test, in presence of 500 μM substrate and 2 mM final PNPA. Controls without NAT or substrate shown no significant activity. Mean values ± SEM are indicated for five independent experiments (n = 5, $^{**}P < 0.01$ and $^{***}P < 0.001$).